

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- 5 (i) APPLICANT:
- (A) NAME: Hoechst Aktiengesellschaft
 - (B) STREET:
 - (C) CITY: Frankfurt
 - (D) FEDERAL STATE: -
 - 10 (E) COUNTRY: Germany
 - (F) POSTAL CODE: 65926
 - (G) TELEPHONE: 069-305-3005
 - (H) TELEFAX: 069-35-7175
 - (I) TELEX: -
- 15 (ii) TITLE OF APPLICATION: Isolation of the genes for
biosynthesizing pseudo-oligosaccharides from
Streptomyces glaucescens GLA.O and their use
- 20 (iii) NUMBER OF SEQUENCES: 13
- (iv) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: floppy disk
 - (B) COMPUTER: IBM PC compatible
 - 25 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version
#1.25 (EPO)
- (2) INFORMATION FOR SEQ ID NO.: 1:
- 30 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - 35 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA

(ix) FEATURES:

(A) NAME/KEY: exon

(B) LOCATION: 1..22

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO.: 1:

CSGGSGSSGC SGGSTTCATS GG

22

10 (2) INFORMATION FOR SEQ ID NO.: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

15 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

20 (ix) FEATURES:

(A) NAME/KEY: exon

(B) LOCATION: 1..24

(xi) SEQUENCE DESCRIPTION: SEQ ID NO.: 2:

25 GGGWVCTGGY VSGGSCCGTA GTTG

24

(2) INFORMATION FOR SEQ ID NO.: 3:

30 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 546 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURES:

(A) NAME/KEY: exon

(B) LOCATION: 1..546

(xi) SEQUENCE DESCRIPTION: SEQ ID NO.: 3:

(x1) SEQUENZBESCHREIBUNG: SEQ ID NO: 3:

CCCCGGCGGG GCGGGGTTC TCGGCTCCGC CTACGTCCGC CGGCTCCTGT CGCCCCGGGC	60
CCCCGGCGGC GTCGCGGTGA CCGTCCTCGA CAAACTCACC TACGCGGCA GCCTCGCCG	120
CCTGCACGCG GTGCGTGACC ATCCCGGCCT CACCTTCGTC CAGGGCGACG TGTGCGACAC	180
CGCGCTCGTC GACACGCTGG CCGCGCGGCA CGACGACATC GTGCACTTCG CGGCCGAGTC	240
GCACGTCGAC CGCTCCATCA CCGACAGCGG TGCTTCACG CGCACCACG TGCTGGGCAC	300
CCAGGTCTTG CTCGACGCCG CGCTCCGCCA CGGTGTGCGC ACCCTCGTGC ACGTCTCCAC	360
CGACGAGGTG TACGGTCCC TCCGACAGG GCGCGCGCG GAGAGCGACC CCCTGCTCCC	420
GACCTCGCCG TACGCGCGT CGAAGGCGGC CTCGACCTC ATGGCGCTCG CCCACCACG	480
CACCCACGGC CTGGACGTCC GGGTGACCCG CTGTTGGAAC AACTACGGCC CGCACCAGTT	540
CCCGGG	546

5 (2) INFORMATION FOR SEQ ID NO.: 4:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 541 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

15 (ix) FEATURES:

- (A) NAME/KEY: exon
(B) LOCATION: 1..541

(x1) SEQUENZBESCHREIBUNG: SEQ ID NO: 4:

CCCCGGGTGC TGGTAGGGGC CGTAGTTGTT GGAGCAGCGG GTGATGCGCA COTCCAGGCC	60
GTGGCTGACG TGCATGGCCA GCGCGAGCAG GTCGCCGAC GCCTTGGAGG TGGCATAGGG	120
GCTGTTGGGG CGCAGCGGCT CGTCCTCCGT CCACGACCCC GTCTCCAGCG AGCCGTAGAC	180
CTCGTCGGTG GACACCTGCA CGAAGGGGGC CACGCCGTGC CGCAGGGCCG CGTCGAGGAG	240
TGTCTGCGTG CCGCGGGCGT TGGTCCGCAC GAACGCGGCG GCATCGAGCA GCGAGCGGTC	300

CACGTGCGAC TCGGCGGCGA GGTGCACGAC CTGGTCCTGG CCGGCCATGA CCCGGTCGAC 360
CAGGTCCGCG TCGCAGATGT CGCCGTGGAC GAAGCGCAGC CGGGGGTGGT CGCGGACCGG 420
GTCGAGGTTG GCGAGGTTGC CGGCGTAGCT CAGGGCGTCG AGCACGGTGA CGACGGCGTC 480
GGGCGGCCCG TCCGGACCGA GGAGGGTGCG GACGTAGTGC GAGCCCATGA ACCCGGCCCG 540
C 541

(2) INFORMATION FOR SEQ ID NO.: 5:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 180 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- 10 (ii) MOLECULE TYPE: protein
- (ix) FEATURES:
(A) NAME/KEY: PCRstrE.Pep
(B) LOCATION: 1..180
- 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO.: 5:

```

Ala Ala Gly Phe Met Gly Ser His Tyr Val Arg Thr Leu Leu Gly Pro
1           5           10           15
Asp Gly Pro Pro Asp Ala Val Val Thr Val Leu Asp Ala Leu Ser Tyr
20           25           30
Ala Gly Asn Leu Ala Asn Leu Asp Pro Val Arg Asp His Pro Arg Leu
35           40           45
Arg Phe Val His Gly Asp Ile Cys Asp Ala Asp Leu Val Asp Arg Val
50           55           60
Met Ala Gly Gln Asp Gln Val Val His Leu Ala Ala Glu Ser His Val
65           70           75           80
Asp Arg Ser Leu Leu Asp Ala Ala Ala Phe Val Arg Thr Asn Ala Gly
85           90           95
Gly Thr Gln Thr Leu Leu Asp Ala Ala Leu Arg His Gly Val Ala Pro
100          105          110
Phe Val Gln Val Ser Thr Asp Glu Val Tyr Gly Ser Leu Glu Thr Gly
115          120          125
Ser Trp Thr Glu Asp Glu Pro Leu Arg Pro Asn Ser Pro Tyr Ala Thr
130          135          140
Ser Lys Ala Ser Gly Asp Leu Leu Ala Leu Ala Met His Val Ser His
145          150          155          160
Gly Leu Asp Val Arg Ile Thr Arg Cys Ser Asn Asn Tyr Gly Pro Tyr
165          170          175
Gln His Pro Gly
180

```

(2) INFORMATION FOR SEQ ID NO.: 6:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 181 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: protein

(ix) FEATURES:

- (A) NAME/KEY: PCR acbD.Pep
 (B) LOCATION: 1..181

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO.: 6:

```

Pro Gly Gly Ala Gly Phe Ile Gly Ser Ala Tyr Val Arg Arg Leu Leu
1           5           10           15
Ser Pro Gly Ala Pro Gly Gly Val Ala Val Thr Val Leu Asp Lys Leu
20           25           30
Thr Tyr Ala Gly Ser Leu Ala Arg Leu His Ala Val Arg Asp His Pro
35           40           45
Gly Leu Thr Phe Val Gln Gly Asp Val Cys Asp Thr Ala Leu Val Asp
50           55           60
Thr Leu Ala Ala Arg His Asp Asp Ile Val His Phe Ala Ala Glu Ser
65           70           75           80
His Val Asp Arg Ser Ile Thr Asp Ser Gly Ala Phe Thr Arg Thr Asn
85           90           95
Val Leu Gly Thr Gln Val Leu Leu Asp Ala Ala Leu Arg His Gly Val
100          105          110
Arg Thr Leu Val His Val Ser Thr Asp Glu Val Tyr Gly Ser Leu Pro
115          120          125
His Gly Ala Ala Ala Glu Ser Asp Pro Leu Leu Pro Thr Ser Pro Tyr
130          135          140
Ala Ala Ser Lys Ala Ala Ser Asp Leu Met Ala Leu Ala His His Arg
145          150          155          160
Thr His Gly Leu Asp Val Arg Val Thr Arg Cys Ser Asn Asn Tyr Gly
165          170          175
Pro His Gln Phe Pro
180

```

(2) INFORMATION FOR SEQ ID NO.: 7:

(i) SEQUENCE CHARACTERISTICS:

5

- (A) LENGTH: 6854 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURES:

- (A) NAME/KEY: "acarbose" biosynthesis gene cluster
- (B) LOCATION: 1..6854

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO.: 7:

CTGCAGGGTT CCCTGGTGCA CGACCCGCCC CTGGTCGACG ACCAGGGCGC TGTCGCAGAT	60
CGCGGCGATG TCGGCGATGT CGTGGCTGGT GAGCACCACG GTGGTGCCCA GTTCCCGGTG	120
GGCGCGGTTG ACCAGCCGGC GCACCGCGTC CTTCAGCACC ATGTCGAGGC CGATCGTGGG	180
CTCGTCCCGA AACAGCACGG CCGGGTCGTG CAGCAGGCTC GCCGCGATCT CGGCGCGCAT	240
GCGCTGTCCG AGGCTGAGCT GCCGCACGGG GGTGGACCCC AGCGCGTCGA TGTCGAGGAG	300
GTCCCGGAAC AGGGCGAGGT TGCGCCCGTA GACCGGTCCG GGGATGTCTG AGATGCGGCG	360
CAGGATGCGG AAGGAGTCGG GTACCGACAG GTCCACCACG AGCTGGCTGC GCTGGCCGAA	420
GACGACGCCG ATCGTGCGGG CGTTGCGCTG CCGGTGCCGG TAGGGCTCCA GCCCGGCGAC	480
CGTGACAGCG CCGGAGGTGG GGGTCATGAT GCCGGTCAGC ATCTTGATCG TGGTCGACTT	540
GCCGGCTCCG TTGGCGCCGA TGTAGGCGGT CTTCTGTCCG GCCGGTATCT CGAAGGAGAC	600
GTCTGTCAGC GCGCGCACGA CGCGGTACCG GCGGGTCAGG AGGGTGGAGA GGCTGCCGAG	660
CAGGCCGGGC TCGCGTTCGG CCAGCCGGAA CTCCTTGACG AGGTGTTCCG CCACGATCAC	720
GCGATCACCC GCTCGACGGC CGTCTCCAGC AGGCGCAGGC CCTCGTCGAG CAGCGCCTCG	780
TCGAGGGTGA ACGGCGGTGC CAGCCGCAGG ATGTGGCCGC CCAGGGAGGT GCGCAGCCCC	840
AGGTCGAGGG CCGTGGTGTA GACGGCCCGG GCGGTCTCGG GGGCGGGTGC CCGGCCGACG	900
GCGTCGGTGA CGAACTCCAG GCCCCACAGC AGTCCGAGGC CGCGTACCTG GCCGAGCTGG	960
GGGAAGCGGG ACTCCAGGGC GCGCAGCCGC TCCTGGATGA GCTCGCCGAG GACGCGCAGC	1020
CGGTGATCA GCCGGTCGCG CTCGACGACC TCCAGCGTGG CGCGGGCGGC GGCGATCCCC	1080
AGTGGGTTGC TCGCGTACGT CGAGGCGTAC GCCCCGGGGT GGCCGCTCC GGCCTGCGCA	1140
GCTTCCGCGC GTCCGGCCAG CACGGCGAAG GGAATCCGC TCGCGGTGCC CTTGGACAGC	1200
ATCGCCAGGT CCGGCTCGAT GCCGAACAGT TCGCTGGCGA GGAAGGCGCC GGTGCGCCCG	1260
CCGCCGTTGA GGACCTCGTC GGCGACGAGC AGCAGCCCGC CGTCCCGGCA GGCGCCGGCG	1320
ATCCGCTCCC AGTAGCCGGG GGGCGGCAGC ATGACGCTG CCGCGCCGAG GACGGGTTCC	1380
AAGACCAGGG CCGAGACGTT GGGCTTCTCC GCGATGTGCC GGCGCACGAG GGTGCGGCAC	1440
CGCACGTCCG ACGAGGGGTA CTCCAGGCCC AGGGGACAGC GGTAGCCAGT AGGGGCTGTA	1500
GCCAGCACGC TGTTGCCGCT GAAGGCCTGG TGCCCGATGT CCCAGTGGAC CAGCATCCGG	1560
GCGCCCATGG TCTTGCCGTG GAAGCCGTGG CGCAGGGCGC AGATCCGGTT GCGGCCCCGC	1620
GCGGCGGTCC CCTGGACGAC CCGCAGGGCG GCCTCGACCA CCTCCGCGCC GGTGGAGAAG	1680
AAGGCGTAGG TGTCGAGCTG TTCGGGCAGC AGCCTGGCGA GCAGTTCAG CAGGCCGGCG	1740
CGGTCCGGCG TGGCGCTGTC GTGGACGTT CACAGGCGGC GGGCCTGGGT GGTGAGTGCC	1800

TCGACGACCT CCGGGTGCCC GTGGCCCACT GACTGGGTGA GGGTCCCGGC CGCGAAGTCG 1860
AGGTACTGGT TGCCGTCCAG GTCGGTCAGA ACGGGACCGC GTCCCTCGGC GAAGACCCGG 1920
CGTCCGTGGA CGGCTTCCTC GGAGGCGCCC GCGCCAGGT GGCGGGCCTC CCGTGCCAGG 1980
TGCTGTGTCT GCCGTAAGCC TGTATCGCT GCCTCTGCTC GTCGGACCGG CTGACGCGAT 2040
CGCCGGCGAA CTGCGTTGTG GCGCACCACG GTTGGGGCGG CTCGGCGCTG AGTCAAACAC 2100
TTGAACACAC ACCGCTGCAA GAGTTTGGG GTTGTTCAG AAAGTTGTTG CGAGCGGCCC 2160
CGGCACTCTG GTTGAGTCGA CGTGCTTACG GCGCCACCAC GCCTCACGTT CGAGGAGGGA 2220
CCTGTGAGAA CAAGCCCGCA GACCGACCGC CTCGCGCGA GGCCGAGGTG AAGGCCCTGG 2280
TCCTGGCAGG TGGAAACCGC AGCAGACTGA GGCCGTTTAC CCACACCGCC GCCAAGCAGC 2340
TGCTCCCCAT CGCCAACAAG CCGTGCTCT TCTACGCGCT GGAGTCCCTC GCCGCGCGCG 2400
GTGTCCGGGA GGCCGGCGTC GTGTTGGCG CGTACGGCCG GGAGATCCGC GAACTCACCG 2460
GCGACGGCAC CGCGTTCCGG TTACGCATCA CCTACCTCCA CCAGCCCCGC CCGCTCGGTC 2520
TCGCGCACGC GGTGCGCATC GCGCGCGCT TCCTGGGCGA CGACGACTTC CTGCTGTACC 2580
TGGGGGACAA CTACCTGCCC CAGGGCGTCA CCGACTTCGC CCGCCAATCG GCCGCCGATC 2640
CCGCGGCGGC CCGGCTGCTG CTCACCCCGG TCGCGGACCC GTCCGCCTTC GGCGTCGCGG 2700
AGGTCGACGC GGACGGGAAC GTGCTGCGCT TGGAGGAGAA ACCCGACGTC CCGCGCAGCT 2760
CGCTCGCGCT CATCGGCGTG TACGCCTTCA GCGCGGCCGT CCACGAGGCG GTACGGGCCA 2820
TCACCCCTC CGCCCGCGGC GAGCTGGAGA TCACCCACGC CGTGCACTGG ATGATCGACC 2880
GGGGCCTGCG CGTACGGGCC GAGACCACCA CCGGGCCCTG GCGCGACACC GGCAGCGCGG 2940
AGGACATGCT GGAGGTCAAC CGTCACGTCC TGGACGGACT GGAGGGCCGC ATCGAGGGGA 3000
AGGTCGACGC GCACAGCAGC CTGGTCGGCC GGGTCCGGGT GGCCGAAGGC GCGATCGTGC 3060
GGGGGTCAAC CGTGGTGGGC CCGGTGGTGA TCGGCGCGGG TGCCGTCGTC AGCAACTCCA 3120
GTGTGCGCCC GTACACCTCC ATCGGGGAGG ACTGCGGGGT CGAGGACAGC GCCATCGAGT 3180
ACTCCGTCTT GCTGCGCGGC GCGCAGGTCG AGGGGGCGTC CCGCATCGAG GCGTCCCTCA 3240
TCGGCCGCGG CGCCGTCGTC GGCCCGGCCC CCGTCTCCC GCAGGCTCAC CGACTGGTGA 3300
TCGGCGACCA CAGCAAGGTG TATCTACCC CATGACCAG ACCATCCTCG TCACCGGCGG 3360
AGCGGGCTTC ATTCGCTCCG CCTACGTCCG CCGGCTCCTG TCGCCCGGGG CCGCCGCGG 3420
CGTCGCGGTG ACCGTCTCG ACAAACCTAC CTACGCGGC AGCCTCGCCC GCTGCACGC 3480
GGTGCGTGAC CATCCCGGCC TCACCTTCGT CCAGGGCGAC GTGTGCGACA CCGCGCTCGT 3540
CGACACGCTG GCGCGCGGC ACGACGACAT CGTGCACTTC GCGGCCGAGT CGCAGCTCGA 3600
CCGCTCCATC ACCGACAGCG GTGCCTTAC CCGCACCAAC GTGCTGGGCA CCCAGGTCTT 3660
GCTCGACGCC GCGCTCCGCC ACGGTGTGCG CACCTTCGTG CACGTCTCCA CCGACGAGGT 3720
GTACGGCTCC CTCGCGCAG GGGCGCGCC GAGAGCGAC CCCCTGCTTC CGACCTCGCC 3780
GTACGCGCGG TCGAAGGCG CCTCGGACCT CATGGCGTTC GCGCACCAAC GCACCCACGG 3840

CCTGGACGTC CGGGTGACCC GCTGTTGAA CAACTTCGGC CCCCACCAGC ATCCCAGAGAA	3900
GCTCATACCG CGCTTCCTGA CCAGCCTCCT GTCCGGCGGC ACCGTTCCCC TCTACGGCGA	3960
CGGGCGGCAC GTGCGCGACT GGCTGCACGT CGACGACCAC GTCAGGGCCG TCGAACTCGT	4020
CCGCGTGTCG GGCCGGCCGG GAGAGATCTA CAACATCGGG GGCGGCACCT CGTGCCCCAA	4080
CCTGGAGCTC ACGCACCGGT TGCTCGCACT GTGCGGCGCG GGCCCGGAGC GCATCGTCCA	4140
CGTCGAGAAC CGCAAGGGGC ACGACCGGCG CTACGCGGTC GACCACAGCA AGATCACCGC	4200
GGAACCTCGGT TACCGGCCGC GCACCGACTT CGCGACCGCG CTGGCCGACA CCGCGAAGTG	4260
GTACGAGCGG CACGAGGACT GGTGGCGTCC CCTGCTCGCC GCGACATGAC GTCGGGCCGG	4320
ACCGCAACCA CCGGCCCCGG CCGGCACACC GCCGCCCGCG GCCGCTGGCC GGCCGGTTCAG	4380
CGTCCGTGAG CCGGGCGCCG GCCGCCCGCG GGGCCGGCGG CGGTGGACCC CCGGACCACC	4440
AGTTCCGGCA TGAAGACGAA TTGGGTGCGC GGCGGGCGCG TTCCGCTCAT CTCCTCCAGC	4500
AGTGCCTCCA CGGCGACCTG CCCCATCGCC TTGACGGGCT GTCTGATGGT GGTCAAGGGA	4560
GGGTCCGTGA AGGCCATGAG CGGCGAGTCG TCGAAGCCGA CCACCGAGAT GTCACCGGGA	4620
ACCGTGAGAC CCGCGCGCGC CGCGGCCCGC ACGGCGCCGA GGGCCATCAT GTCGCTGGCG	4680
CACATGACGG CGGTGCAGCC CAGGTCGATC AGCGCGGACG CGGCGGCCTG GCCCCCTCC	4740
AGGGAGAACA GCGAGTGCTG CACGAGCTCC TCGGACTCCC GCGCCGACAC TCCCAGGTGC	4800
TCCCGCACGC CGGCCCGGAA CCCCTCGATC TTCCGCTGCA CCGGCACGAA GCGGGCGGGC	4860
CCGACGGCGA GGCCGACGCG CTCGTGCCCC AGCTCCGCCA GGTGCGCCAC GGCCAGGCGC	4920
ATCGCGGCCC GGTCTGTCGG GGAGACGAAG GGTGCCTCGA TCCGGGGCGA GAACCCGTTT	4980
ACGAGGACGA AGGGCACCTG CCGCTCGTGC AGCGGGCCGT ACCGTCCGGT CTCGGCGGTG	5040
GTGTCCCGGT GCAGTCCGGA GACGAAGATG ATGCCGACA CCGCGCGGTC CACGAGCATC	5100
TCCGTGAGTT CGTCTCGGT CGAGCCGCCC GGGGTCTGCG TGGCGAGCAC GGGCGTGTAG	5160
CCCTGACGCG TGAGCGCCTG CCCCATCACC TGGGCCAGTG CCGGGAAGAA GGGGTGTGTC	5220
AGTTCGGGGG TGACCAGTCC GACCAGCTCG GCGCGGCGCT GTGCGCGCCG CTGCTCGTAG	5280
CCCAGCGCGT CCAGTGCAGT CAGCACCGAG TCGCGGGTGC CGGTGGCCAC ACCGCGCGCA	5340
CCGTTGAGCA CCGGCTGAC CGTGGCCTTG CTGACGCCCC CCGGGGCTGC GATGTGCGCG	5400
AGCCGCATGG TCATGGCAAC GCACTCTACC TGTCGGGGCG TCAGGGCGTG CCCACCGCGC	5460
GCGGAACCGG CGGACTGCGG GGCACGGCCC GTCCGCCGCC CACGGACCAC GCGCCCCAAA	5520
CGATGGCTGA AAATGCTTGC AGCAAAATTG CGCAACGTCT TTCGGCGGCT TTTCGATCCT	5580
GTTACGTTCC TGGCAACCCC GCGCGCGCGC AGAAGCGGTT GGCGTGAGGC GTCCAGACCT	5640
CCGCCCCATT CCGGGATCAC TCAGGGGAGT TCACAATGCG GCGTGGCATT GCGGCCACCG	5700
CGCTGTTGCG GGCTGTGGCC ATGACGGCAT CGGCGTGTGG CCGGGGCGAC AACGGCGGAA	5760
GCGGTACCGA CCGGGCGGC ACGGAGCTGT CCGGGACCGT CACCTTCTGG GACACGTCCA	5820
ACGAAGCCGA GAAGGCGACG TACCAGGCCC TCGCGGAGGG CTTCGAGAAG GAGCACCCGA	5880

AGGTCGACGT CAAGTACGTC AACGTCCCGT TCGGCGAGGC GAACGCCAAG TTCAAGAACG 5940
CCGCGGGCGG CAACTCCGGT GCGCCGGACG TGATGCGTAC GGAGGTCGCC TGGGTCGCGG 6000
ACTTCGCCAG CATCGGCTAC CTCGCCCCGC TCGACGGCAC GCGCGCCCTC GACGACGGGT 6060
CGGACCACCT TCCCCAGGGC GGCAGACCA GGTACGAGGG GAAGACCTAC GCGGTCCCCG 6120
AGGTGATCGA CACCCTGGCG CTCTTCTACA ACAAGGAACT GCTGACGAAG GCCGGTGTCTG 6180
AGGTGCCGGG CTCCCTCGCC GAGCTGAAGA CGGCCGCCGC CGAGATCACC GAGAAGACCG 6240
GCGCGAGCGG CCTCTACTGC GGGGCGACGA CCGTACTTGT GTTCCTGCCC TACCTCTACG 6300
GGGAGGGCGG CGACCTGGTC GACGAGAAGA ACAAGACCGT CACGGTCGAC GACGAAGCCG 6360
GTGTGCGCGC CTACCGCGTC ATCAAGGACC TCGTGGACAG CAAGGCGGCC ATCACCAGC 6420
CGTCCGACGG CTGGAACAAC ATGCAGAAGC CCTTCAAGTC GGGCAAGGTC GCCATGATGG 6480
TCAACGGCCC CTGGGCCATC GAGGACGTCA AGGCGGGAGC CCGCTTCAAG GACGCCGGCA 6540
ACCTGGGGGT CGCCCCGTC CCGGCCGGCA GTGCCGACA GGGCTCTCCC CAGGGCGGGT 6600
GGAACCTCTC GGTGTACGCG GGCTCGAAGA ACCTCGACGC CTCCTACGCC TTCGTGAAGT 6660
ACATGAGCTC CGCCAAGGTG CAGCAGCAGA CCACCGAGAA GCTGAGCCTG CTGCCCCACC 6720
GCACGTCCGT CTACGAGGTC CCGTCCGTCG CGGACAACGA GATGGTGAAG TTCTTCAAGC 6780
CGGCCGTCTA CAAGGCCGTC GAACGGCCGT GGATCGCCGA GGGCAATGCC CTCTTCGAGC 6840
CGATCCGGCT GCAG 6854

(2) INFORMATION FOR SEQ ID NO.: 8:

(i) SEQUENCE CHARACTERISTICS:

5

- (A) LENGTH: 240 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: protein

(ix) FEATURES:

- (A) NAME/KEY: acbA
- (B) LOCATION: 1..240

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO.: 8:

Val Ile Val Ala Glu His Leu Val Lys Glu Phe Arg Leu Ala Glu Arg
1 5 10 15
Glu Pro Gly Leu Leu Gly Ser Leu Ser Thr Leu Leu Thr Arg Arg Tyr
20 25 30
Arg Val Val Arg Ala Val Asp Asp Val Ser Phe Glu Ile Pro Ala Gly
35 40 45
Thr Lys Thr Ala Tyr Ile Gly Ala Asn Gly Ala Gly Lys Ser Thr Thr
50 55 60
Ile Lys Met Leu Thr Gly Ile Met Thr Pro Thr Ser Gly Arg Cys Thr
65 70 75 80
Val Ala Gly Leu Glu Pro Tyr Arg His Arg Gln Arg Asn Ala Arg Thr
85 90 95
Ile Gly Val Val Phe Gly Gln Arg Ser Gln Leu Trp Trp Asp Leu Ser
100 105 110
Val Pro Asp Ser Phe Arg Ile Leu Arg Arg Ile Tyr Asp Ile Pro Gly
115 120 125
Pro Val Tyr Arg Arg Asn Leu Ala Leu Phe Arg Asp Leu Leu Asp Ile
130 135 140
Asp Ala Leu Gly Ser Thr Pro Val Arg Gln Leu Ser Leu Gly Gln Arg
145 150 155 160
Met Arg Ala Glu Ile Ala Ala Ser Leu Leu His Asp Pro Ala Val Leu
165 170 175
Phe Trp Asp Glu Pro Thr Ile Gly Leu Asp Met Val Leu Lys Asp Ala
180 185 190
Val Arg Arg Leu Val Asn Arg Ala His Arg Glu Leu Gly Thr Thr Val
195 200 205
Val Leu Thr Ser His Asp Ile Ala Asp Ile Ala Ala Ile Cys Asp Ser
210 215 220
Ala Leu Val Val Asp Gln Gly Arg Val Val His Gln Gly Thr Leu Gln
225 230 235 240

(2) INFORMATION FOR SEQ ID NO.: 9:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 429 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10

- (ii) MOLECULE TYPE: protein

- (ix) FEATURES:

15

- (A) NAME/KEY: acbB
(B) LOCATION: 1..429

(xi) SEQUENCE DESCRIPTION: SEQ ID NO.: 9:

Met	Thr	Gly	Leu	Arg	Gln	Thr	Gln	His	Leu	Ala	Arg	Glu	Ala	Arg	His
1				5					10					15	
Leu	Ala	Pro	Gly	Ala	Ser	Glu	Glu	Ala	Val	His	Gly	Arg	Arg	Val	Phe
			20					25					30		
Ala	Glu	Gly	Arg	Gly	Pro	Val	Leu	Thr	Asp	Leu	Asp	Gly	Asn	Gln	Tyr
		35					40					45			
Leu	Asp	Phe	Ala	Ala	Gly	Thr	Leu	Thr	Gln	Ser	Leu	Gly	His	Gly	His
	50					55					60				
Pro	Glu	Val	Val	Glu	Ala	Leu	Thr	Thr	Gln	Ala	Arg	Arg	Leu	Trp	Asn
65					70					75					80
Val	His	Asp	Ser	Ala	Thr	Pro	Asp	Arg	Ala	Gly	Leu	Leu	Glu	Leu	Leu
				85					90					95	

Ala Arg Leu Leu Pro Glu Gln Leu Asp Thr Tyr Ala Phe Phe Ser Thr
 100 105 110
 Gly Ala Glu Val Val Glu Ala Ala Leu Arg Val Val Gln Ala Thr Ala
 115 120 125
 Ala Pro Gly Arg Asn Arg Ile Cys Ala Leu Arg His Gly Phe His Gly
 130 135 140
 Lys Thr Met Gly Ala Arg Met Leu Val His Trp Asp Ile Gly His Gln
 145 150 155 160
 Ala Phe Ser Gly Asn Ser Val Leu Ala Thr Ala Pro Thr Gly Tyr Arg
 165 170 175
 Cys Pro Leu Gly Leu Glu Tyr Pro Ser Cys Asp Val Arg Cys Ala Thr
 180 185 190
 Leu Val Arg Arg His Ile Ala Glu Lys Pro Asn Val Ser Ala Leu Val
 195 200 205
 Phe Glu Pro Val Leu Gly Ala Ala Gly Val Ile Val Pro Pro Pro Gly
 210 215 220
 Tyr Trp Glu Arg Ile Ala Gly Ala Cys Arg Asp Gly Gly Val Leu Leu
 225 230 235 240
 Val Ala Asp Glu Val Leu Thr Gly Gly Gly Arg Thr Gly Ala Phe Leu
 245 250 255
 Ala Ser Glu Leu Phe Gly Ile Glu Pro Asp Leu Ala Met Leu Ser Lys
 260 265 270
 Gly Thr Ala Ser Gly Phe Pro Phe Ala Val Leu Ala Gly Arg Ala Glu
 275 280 285
 Ala Ala Gln Ala Gly Gly Gly His Pro Gly Ala Tyr Ala Ser Thr Tyr
 290 295 300
 Ala Ser Asn Pro Leu Gly Ile Ala Ala Ala Arg Ala Thr Leu Glu Val
 305 310 315 320
 Val Glu Arg Asp Arg Leu Ile Asp Arg Val Arg Val Leu Gly Glu Leu
 325 330 335
 Ile Gln Glu Arg Leu Arg Ala Leu Glu Ser Arg Phe Pro Gln Leu Gly
 340 345 350
 Gln Val Arg Gly Leu Gly Leu Leu Trp Gly Leu Glu Phe Val Thr Asp
 355 360 365
 Ala Val Gly Arg Ala Pro Ala Pro Glu Thr Ala Arg Ala Val Tyr Thr
 370 375 380
 Thr Ala Leu Asp Leu Gly Leu Arg Thr Ser Leu Gly Gly His Ile Leu
 385 390 395 400
 Arg Leu Ala Pro Pro Phe Thr Leu Asp Glu Ala Leu Leu Asp Glu Gly
 405 410 415
 Leu Arg Leu Leu Glu Thr Ala Val Glu Arg Val Ile Ala
 420 425

(2) INFORMATION FOR SEQ ID NO.: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 355 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

- (ii) MOLECULE TYPE: protein

- (ix) FEATURES:

10

- (A) NAME/KEY: acbC
- (B) LOCATION: 1..355

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO.: 10:

Val Lys Ala Leu Val Leu Ala Gly Gly Thr Gly Ser Arg Leu Arg Pro
 1 5 10 15
 Phe Thr His Thr Ala Ala Lys Gln Leu Leu Pro Ile Ala Asn Lys Pro
 20 25 30
 Val Leu Phe Tyr Ala Leu Glu Ser Leu Ala Ala Ala Gly Val Arg Glu
 35 40 45
 Ala Gly Val Val Val Gly Ala Tyr Gly Arg Glu Ile Arg Glu Leu Thr
 50 55 60
 Gly Asp Gly Thr Ala Phe Gly Leu Arg Ile Thr Tyr Leu His Gln Pro
 65 70 75 80
 Arg Pro Leu Gly Leu Ala His Ala Val Arg Ile Ala Arg Gly Phe Leu
 85 90 95
 Gly Asp Asp Asp Phe Leu Leu Tyr Leu Gly Asp Asn Tyr Leu Pro Gln
 100 105 110
 Gly Val Thr Asp Phe Ala Arg Gln Ser Ala Ala Asp Pro Ala Ala Ala
 115 120 125
 Arg Leu Leu Leu Thr Pro Val Ala Asp Pro Ser Ala Phe Gly Val Ala
 130 135 140
 Glu Val Asp Ala Asp Gly Asn Val Leu Arg Leu Glu Glu Lys Pro Asp
 145 150 155 160
 Val Pro Arg Ser Ser Leu Ala Leu Ile Gly Val Tyr Ala Phe Ser Pro
 165 170 175
 Ala Val His Glu Ala Val Arg Ala Ile Thr Pro Ser Ala Arg Gly Glu
 180 185 190
 Leu Glu Ile Thr His Ala Val Gln Trp Met Ile Asp Arg Gly Leu Arg
 195 200 205
 Val Arg Ala Glu Thr Thr Thr Arg Pro Trp Arg Asp Thr Gly Ser Ala
 210 215 220
 Glu Asp Met Leu Glu Val Asn Arg His Val Leu Asp Gly Leu Glu Gly
 225 230 235 240
 Arg Ile Glu Gly Lys Val Asp Ala His Ser Thr Leu Val Gly Arg Val
 245 250 255
 Arg Val Ala Glu Gly Ala Ile Val Arg Gly Ser His Val Val Gly Pro
 260 265 270
 Val Val Ile Gly Ala Gly Ala Val Val Ser Asn Ser Ser Val Gly Pro
 275 280 285
 Tyr Thr Ser Ile Gly Glu Asp Cys Arg Val Glu Asp Ser Ala Ile Glu
 290 295 300
 Tyr Ser Val Leu Leu Arg Gly Ala Gln Val Glu Gly Ala Ser Arg Ile
 305 310 315 320
 Glu Ala Ser Leu Ile Gly Arg Gly Ala Val Val Gly Pro Ala Pro Arg
 325 330 335
 Leu Pro Gln Ala His Arg Leu Val Ile Gly Asp His Ser Lys Val Tyr
 340 345 350
 Leu Thr Pro
 355

(2) INFORMATION FOR SEQ ID NO.: 11:

(i) SEQUENCE CHARACTERISTICS:

5

(A) LENGTH: 325 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

10

(ix) FEATURES:

(A) NAME/KEY: acbD

(B) LOCATION: 1..325

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO.: 11:

Met	Thr	Thr	Thr	Ile	Leu	Val	Thr	Gly	Gly	Ala	Gly	Phe	Ile	Arg	Ser
1				5					10					15	
Ala	Tyr	Val	Arg	Arg	Leu	Leu	Ser	Pro	Gly	Ala	Pro	Gly	Gly	Val	Ala
			20					25					30		
Val	Thr	Val	Leu	Asp	Lys	Leu	Thr	Tyr	Ala	Gly	Ser	Leu	Ala	Arg	Leu
		35					40					45			
His	Ala	Val	Arg	Asp	His	Pro	Gly	Leu	Thr	Phe	Val	Gln	Gly	Asp	Val
	50					55					60				
Cys	Asp	Thr	Ala	Leu	Val	Asp	Thr	Leu	Ala	Ala	Arg	His	Asp	Asp	Ile
65					70					75					80
Val	His	Phe	Ala	Ala	Glu	Ser	His	Val	Asp	Arg	Ser	Ile	Thr	Asp	Ser
			85						90					95	
Gly	Ala	Phe	Thr	Arg	Thr	Asn	Val	Leu	Gly	Thr	Gln	Val	Leu	Leu	Asp
			100					105					110		
Ala	Ala	Leu	Arg	His	Gly	Val	Arg	Thr	Phe	Val	His	Val	Ser	Thr	Asp
		115					120					125			
Glu	Val	Tyr	Gly	Ser	Leu	Pro	His	Gly	Ala	Ala	Ala	Glu	Ser	Asp	Pro
	130					135					140				


```

Leu Leu Pro Thr Ser Pro Tyr Ala Ala Ser Lys Ala Ala Ser Asp Leu
145                      150                      155                      160
Met Ala Leu Ala His His Arg Thr His Gly Leu Asp Val Arg Val Thr
                      165                      170                      175
Arg Cys Ser Asn Asn Phe Gly Pro His Gln His Pro Glu Lys Leu Ile
                      180                      185                      190
Pro Arg Phe Leu Thr Ser Leu Leu Ser Gly Gly Thr Val Pro Leu Tyr
                      195                      200                      205
Gly Asp Gly Arg His Val Arg Asp Trp Leu His Val Asp Asp His Val
210                      215                      220
Arg Ala Val Glu Leu Val Arg Val Ser Gly Arg Pro Gly Glu Ile Tyr
225                      230                      235                      240
Asn Ile Gly Gly Gly Thr Ser Leu Pro Asn Leu Glu Leu Thr His Arg
                      245                      250                      255
Leu Leu Ala Leu Cys Gly Ala Gly Pro Glu Arg Ile Val His Val Glu
                      260                      265                      270
Asn Arg Lys Gly His Asp Arg Arg Tyr Ala Val Asp His Ser Lys Ile
                      275                      280                      285
Thr Ala Glu Leu Gly Tyr Arg Pro Arg Thr Asp Phe Ala Thr Ala Leu
290                      295                      300
Ala Asp Thr Ala Lys Trp Tyr Glu Arg His Glu Asp Trp Trp Arg Pro
305                      310                      315                      320
Leu Leu Ala Ala Thr
                      325

```

(2) INFORMATION FOR SEQ ID NO.: 12:

(i) SEQUENCE CHARACTERISTICS:

5

- (A) LENGTH: 345 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: protein

(ix) FEATURES:

- (A) NAME/KEY: acbE
- (B) LOCATION: 1..345

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO.: 12:

Met Thr Met Arg Leu Ala Asp Ile Ala Ala Arg Ala Gly Val Ser Lys
 1 5 10 15
 Ala Thr Val Ser Arg Val Leu Asn Gly Ala Arg Gly Val Ala Thr Gly
 20 25 30
 Thr Arg Asp Ser Val Leu Thr Ala Leu Asp Ala Leu Gly Tyr Glu Gln
 35 40 45
 Pro Ala Arg Gln Arg Arg Ala Glu Leu Val Gly Leu Val Thr Pro Glu
 50 55 60
 Leu Asp Asn Pro Phe Phe Pro Ala Leu Ala Gln Val Met Gly Gln Ala
 65 70 75 80
 Leu Thr Arg Gln Gly Tyr Thr Pro Val Leu Ala Thr Gln Thr Pro Gly
 85 90 95
 Gly Ser Thr Glu Asp Glu Leu Thr Glu Met Leu Val Asp Arg Gly Val
 100 105 110
 Ser Gly Ile Ile Phe Val Ser Gly Leu His Ala Asp Thr Thr Ala Glu
 115 120 125
 Thr Gly Arg Tyr Gly Arg Leu His Glu Arg Gln Val Pro Phe Val Leu
 130 135 140
 Val Asn Gly Phe Ser Pro Arg Ile Glu Ala Pro Phe Val Ser Pro Asp
 145 150 155 160
 Asp Arg Ala Ala Met Arg Leu Ala Val Ala His Leu Ala Glu Leu Gly
 165 170 175
 His Glu Arg Val Gly Leu Ala Val Gly Pro Ala Arg Phe Val Pro Val
 180 185 190
 Gln Arg Lys Ile Glu Gly Phe Arg Ala Gly Val Arg Glu His Leu Gly
 195 200 205
 Val Ser Ala Arg Glu Ser Glu Glu Leu Val Gln His Ser Leu Phe Ser
 210 215 220
 Leu Glu Gly Gly Gln Ala Ala Ala Ser Ala Leu Ile Asp Leu Gly Cys
 225 230 235 240
 Thr Ala Val Met Cys Ala Ser Asp Met Met Ala Leu Gly Ala Val Arg
 245 250 255
 Ala Ala Arg Arg Arg Gly Leu Thr Val Pro Gly Asp Ile Ser Val Val
 260 265 270
 Gly Phe Asp Asp Ser Pro Leu Met Ala Phe Thr Asp Pro Pro Leu Thr
 275 280 285
 Thr Ile Arg Gln Pro Val Lys Ala Met Gly Gln Val Ala Val Asp Ala
 290 295 300
 Leu Leu Glu Glu Met Ser Gly Thr Pro Pro Pro Arg Thr Glu Phe Val
 305 310 315 320
 Phe Met Pro Glu Leu Val Val Arg Gly Ser Thr Ala Ala Gly Pro Arg
 325 330 335
 Gly Gly Arg Arg Pro Ala His Gly Arg
 340 345

(2) INFORMATION FOR SEQ ID NO.: 13:

(i) SEQUENCE CHARACTERISTICS:

5

(A) LENGTH: 393 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

(ii) MOLECULE TYPE: protein

(ix) FEATURES:

(A) NAME/KEY: acbF

(B) LOCATION: 1..393

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO.: 13:

Met Arg Arg Gly Ile Ala Ala Thr Ala Leu Phe Ala Ala Val Ala Met
 1 5 10 15
 Thr Ala Ser Ala Cys Gly Gly Gly Asp Asn Gly Gly Ser Gly Thr Asp
 20 25 30
 Ala Gly Gly Thr Glu Leu Ser Gly Thr Val Thr Phe Trp Asp Thr Ser
 35 40 45
 Asn Glu Ala Glu Lys Ala Thr Tyr Gln Ala Leu Ala Glu Gly Phe Glu
 50 55 60
 Lys Glu His Pro Lys Val Asp Val Lys Tyr Val Asn Val Pro Phe Gly
 65 70 75 80
 Glu Ala Asn Ala Lys Phe Lys Asn Ala Ala Gly Gly Asn Ser Gly Ala
 85 90 95
 Pro Asp Val Met Arg Thr Glu Val Ala Trp Val Ala Asp Phe Ala Ser
 100 105 110
 Ile Gly Tyr Leu Ala Pro Leu Asp Gly Thr Pro Ala Leu Asp Asp Gly
 115 120 125
 Ser Asp His Leu Pro Gln Gly Gly Ser Thr Arg Tyr Glu Gly Lys Thr
 130 135 140
 Tyr Ala Val Pro Gln Val Ile Asp Thr Leu Ala Leu Phe Tyr Asn Lys
 145 150 155 160
 Glu Leu Leu Thr Lys Ala Gly Val Glu Val Pro Gly Ser Leu Ala Glu
 165 170 175
 Leu Lys Thr Ala Ala Ala Glu Ile Thr Glu Lys Thr Gly Ala Ser Gly
 180 185 190
 Leu Tyr Cys Gly Ala Thr Thr Arg Thr Trp Phe Leu Pro Tyr Leu Tyr
 195 200 205
 Gly Glu Gly Gly Asp Leu Val Asp Glu Lys Asn Lys Thr Val Thr Val
 210 215 220
 Asp Asp Glu Ala Gly Val Arg Ala Tyr Arg Val Ile Lys Asp Leu Val
 225 230 235 240
 Asp Ser Lys Ala Ala Ile Thr Asp Ala Ser Asp Gly Trp Asn Asn Met
 245 250 255
 Gln Asn Ala Phe Lys Ser Gly Lys Val Ala Met Met Val Asn Gly Pro
 260 265 270
 Trp Ala Ile Glu Asp Val Lys Ala Gly Ala Arg Phe Lys Asp Ala Gly
 275 280 285
 Asn Leu Gly Val Ala Pro Val Pro Ala Gly Ser Ala Gly Gln Gly Ser
 290 295 300
 Pro Gln Gly Gly Trp Asn Leu Ser Val Tyr Ala Gly Ser Lys Asn Leu
 305 310 315 320

Asp Ala Ser Tyr Ala Phe Val Lys Tyr Met Ser Ser Ala Lys Val Gln
325 330 335

Gln Gln Thr Thr Glu Lys Leu Ser Leu Leu Pro Thr Arg Thr Ser Val
340 345 350

Tyr Glu Val Pro Ser Val Ala Asp Asn Glu Met Val Lys Phe Phe Lys
355 360 365

Pro Ala Val Asp Lys Ala Val Glu Arg Pro Trp Ile Ala Glu Gly Asn
370 375 380

Ala Leu Phe Glu Pro Ile Arg Leu Gln
385 390